UNIVERSITY OF JOHANNESBURG

DEPARTMENT OF ECONOMICS AND ECONOMETRICS

AUCKLAND PARK KINGSWAY CAMPUS

SECOND SEMESTER EXAMINATION (2016)

Module: Econometrics 4B (EKN4815)

Examiner: Dr Kwame Osei-Assibey

Examination Duration: 3 Hours

Marks: 100

Instructions

- **All sections and questions are compulsory.**
- **No examination paper are to be removed from the room**

Section A (Theory and Concepts)

Question One **[7]**

a. Briefly discuss three benefits of using panel data rather than pure cross-sections or time series datasets. **(3 Marks)**

b. Explain the term "endogenous variable" as it applies to econometric analysis. Mention three different potential sources of endogeneity with which an econometrician needs to be concerned about. **(4 Marks)**

Question Two **[9]**

Consider the following regression model;

$y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots \beta_k x_{itk} + \underbrace{a_i + u_t}_{v_{it}}$:Where v_{it} is the composite error term;

μ_{it} is the idiosyncratic error term; α_i is the unobserved heterogeneity component.

a. You are thinking about using either fixed effect or random effect technique to analyse a panel data based on the regression model above: State what technique is appropriate under each of the following conditions.

i. If Cov (x_{it}, α_i)≠0
ii. If Cov (x_{it}, α_i)=0

(2 Marks)

b. You are presented with data and has been asked to construct the random effect estimator for the given model and the data. The software you are using can only perform OLS estimation but not the random effect estimation. Luckily for you, a friend whispered to you that $\sigma_a^2 = E(\alpha_i^2)$ and $\sigma_u^2 = E(u_t^2)$. Use this information to construct a random-effects estimator for your data and model. **(3 Marks)**

c. If $\sigma_a^2 \to 0$ what value does the random effect transformation parameter tend to? Which estimator does it become? **(2 Marks)**

d. If $\sigma_u^2 \to 0$ what value does the random effect transformation parameter tend to? Which estimator does it become? **(2 Marks)**

Question Three [12]

A student identifies four explanatory variables that he believes can be used to explain changes in y_1 (the dependent variable). He discussed his regression model with his friend who advised him that one of the explanatory variables (x_1)is endogenous and two (x_2 and x_4) are potential valid instruments for x_1. The student is interested in performing an IV regression using the 2SLS. The only problem he faces is that the software he is using can only performs OLS estimation. Explain to him how he can perform the 2SLS process 'manually'. In your explanation, be sure to also explain the following

i. What we mean by relevance and weakness of the instruments and how to test them.
ii. The standard errors of β_{IV} from 2SLS performed 'manually' compared to the standard errors from 2SLS automatically estimated by STATA.

(12 Marks)

Question Four [12]

a. What is the difference between a probit and a logit model? On what basis can you choose which one to use, and why does it matter? **(5 Marks)**
b. This hypothetical question looks at how education, age and husband's income are related to a married woman labour supply :
   i. Construct a latent variable for a probit model on the probability of employment, with education (in years), age, and age squared as explanatory variables. Give an interpretation of the latent variable. **(5 Marks)**
   ii. What is the partial effect of education on the employment probability? How is this different from partial effect of the average **(2 Marks)**

Section B (Practice and Applications Using STATA)

Question Five **[25]**

The following question is based on the CPS78_85 dataset. Among other analysis that can be done with this data, one can analyse the wage differentials between different races for the years 1978 and 1975 (note that the years are recorded in the data as 78 and 85). This task STRICTLY requires you to analyse the impact of been non-white on wages (log of wages).

a. Write down the functional form equation to be estimated. **(3 Marks)**
b. Use STATA to perform **separate** OLS regressions for the years 1978 and 1985 and present your estimated models with t-statistics in brackets right below the estimated coefficients. Discuss the statistical significance of the estimated coefficients. **(8 Marks)**
c. Explain what the values of the intercepts in both regressions represent. **(3 Marks)**
d. Explain what the values of the coefficients in both regressions represent. **(3 Marks)**
e. What is the marginal effect of being non-white on wages between 1978 and 1985?
f. By using STATA and pooling data for the two years, test for the significance of the marginal effect you calculated in (e) above. **(8 Marks)**

Question Six **[35]**

This exercise uses the Mroz dataset. We want to analyse the impact of work experience (exper), work experience squared (expersq), woman's education (educ), woman's age (age), the family income excluding the wife's wage (nwifeinc), number of kids less than six years of age (kidslt6) and number of kids greater than or equal to six years of age (kidsge6) on wages (lwage) of working married women.

a. Estimate the model using OLS and present your estimated model with t-statistics in brackets right below the estimated coefficient. Discuss the statistical significance of the estimated coefficients and explain whether your obtained results are consistent with your expectations. **(8 Marks)**
b. Indicate two reasons why the OLS estimation might produce a biased coefficient on educ. **(2 Marks)**
c. Consider fatheduc, motheduc and huseduc as potential instruments for women years of schooling ; list two reasons why you think these variables can serve as good instruments. **(2 Marks)**

d. Perform the necessary test(s) to ascertain if these instruments are relevant and strong for educ. Present all the statistics you used to arrive at your conclusion. **(5 Marks)**
e. Using fatheduc, motheduc and huseduc as instruments perform an IV regression. Present your estimated model with t-statistics in brackets right under the estimated coefficient. Discuss the statistical significance of the estimated educ coefficient. Did the inclusion of the other variables improve your estimation? Explain. **(8 Marks)**
f. Test for endogeneity. Based on the evidence from the endogeneity test results, would you say it was wise afterall to perform the IV regression? Explain. **(5 Marks)**
g. Test to ascertain if there is an over-identification problem. **(5 Marks)**

END